REMARKS

Claims 42, 43 and 45 were pending in the case. Upon entry of this Amendment, Claim 45 will be cancelled, and Claims 42 and 43 will remain pending in the case.

I. The Sequence Listing

The Office Action objects to the Sequence Listing provided with the case. It alleges that at page 14, lines 18-19 nucleic acid sequences are included which are in excess of 10 bases, but which lack any Seq. ID. No. It adds that the sequences are similar, but not identical to Seq. ID. Nos. 3 and 4, but differ at page 14 in that they comprise inosine residues. The Office Action states that the sequences must be disclosed in the Sequence Listing, and that Applicants must provide a new sequence listing in paper and computer readable form and a statement that the two are the same and that no new matter is introduced.

In response, Applicants provide a new Sequence Listing that Applicants believe will overcome the rejections raised by the Office Action. It is provided in both paper and computer readable form, and the requested Statement of identity and no new matter is also included herewith. Applicants believe the foregoing will overcome the Office Action's objection to the Sequence Listing.

II. The Objections to Claims 42 and 43

The Office Action objects to Claim 42, alleging that as amended, it is drawn to non-elected subject matter. The Office Action alleges that Applicants elected methods of screening compounds which alter the conductive properties of acetylcholine receptors. The Office Action adds that Claim 42 was amended to encompass methods of screening compounds that alter at least one property of an acetylcholine receptor, and therefore embraces receptor properties other than conductivity, and is therefore alleged by the Office Action to be beyond the scope of the elected invention.

The Office Action objects to Claim 43, alleging that by its amendment it therefore embraces non-elected subject matter, but does not specify with

particularity how Claim 43, as previously amended, extends beyond the scope of the elected material.

By way of response, Applicants have amended Claim 42 to limit it to the conductive property as requested by the Office Action.

Applicants are not sure why Claim 43 was believed by the Office Action to encompass non-elected subject matter, but respectfully assert that Claim 43 as amended by this Amendment is directed to elected subject matter. Applicants would appreciate discussing Claim 43 with the Examiner in a telephone conversation, if Claim 43 as amended herein is not deemed allowable, before the Examiner issues another Office Action with respect to Claim 43.

Applicants believe the objections to Claims 42 and 43 have been overcome, and review and reconsideration of the claims are respectfully requested.

III. The Section 112, First Paragraph Rejections - Written Description

Claims 42, 43 and 45 stand rejected under 35 U.S.C. Section 112, first paragraph, for the reasons of record in paper number 25. Applicants were unable to locate a paper no. 25, and respectfully request the Examiner to identify paper number 25 with greater particularity.

The Office Action goes on to state that Claims 42 and 45 are drawn to methods of screening compounds with alter the conductive properties of acetylcholine receptors, and that Claim 43 is drawn to a method of identifying compounds that bind to an acetylcholine receptor. The Office Action adds that the methods employ nucleic acids from the group listed on page 5 of the Office Action, and that embodiments of Claim 45 employ polypeptides encoded by the nucleic acids set forth above.

The Office Action adds that the first of these groups of nucleic acids (Applicants assume the Office Action refers to nucleic acids consisting of Seq. ID. No. 1) which are fully disclosed and adequately described, which Applicants acknowledge with appreciation.

The Office Action adds that similarly, the polypeptide of Seq. ID. No. 2 is fully disclosed and adequately described, which Applicants also acknowledge with appreciation.

The Office Action alleges that the remaining groups of sequences represent genuses,

The Office Action alleges that the genus in not enabled, alleging, in summary, that there is only a disclosure of a single member (e.g. SEQ ID NO.:2) of the claimed genuses, nor is there any relevant identifying characteristic described, such as a correlation between a specific structure and the required function.

By way of response, Applicants have cancelled Claim 45, and have amended Claim 42 to limit it to the conductive property and have otherwise amended Claims 42 and 43 to address the issues raised by this section of the Office Action.

Applicants believe amended Claims 42 and 43 overcome these objections/rejections and are in condition for allowance. Review and reconsideration of Claims 42 and 43 as amended herein is respectfully requested.

IV. The Section 112, First Paragraph Rejections - Enablement

Claims 42, 43 and 45 stand rejected under 35 U.S.C. Section 112, first paragraph, on the basis that the specification, while enabling for methods of screening compounds which alter the conductive properties of acetylcholine receptors, or which bind to acetylcholine receptors as listed in the Office Action, allegedly does not reasonably provide enablement for such methods:

that employ subsequences of SEQ ID NO:1 that are only 14 bases in length, or

that employ nucleic acid sequences encoding polypeptides that differ from the entire sequence of SEQ ID NO:2 or from subsequences of SEQ ID NO:2, or

for methods that employ variants of SEQ ID NO:2 or variants of its subsequences.

The Office Action alleges that the specification does not enable compounds useful for either or both of crop protection and pharmaceutical treatment of humans. The Office Action alleges that the specification does not enable any person skilled in the art to which it pertains, or to with which it is most closely connected, to make the invention commensurate in scope with these claims.

The Office Action then provides a discussion of SEQ. ID NO:1, and concludes that it is unrealistic to assume that a nucleic acid of only 14 nucleotides could encode a polypeptide with the claimed function, nor could its complement.

The Office Action then discusses nucleic acids that encode variants of SEQ ID NO:2 or its subsequences, concluding that the specification is not enabling because it does not teach which variants will provide a functional beta subunit and which will not. The Office Action, in summary, concludes that determining this will require undue experimentation.

The Office Action also alleges the specification fails to provide an enabling disclosure because it fails to teach what is useful for crop protection and/or pharmaceutical treatment of humans.

The Office Action also alleges that the specification fails to provide an enabling disclosure because essential elements of the invention are incorporated by reference to a non-patent publication in its reference to the use of GCG program GAP, Version 10.0 using "standard settings".

Claims 42, 43 and 45 are rejected under 35 U.S.C. Section 112, second paragraph, with the Office Action alleging that the claims are indefinite because they recite "the biological function of an acetylcholine receptor" without antecedent basis. The Office Action adds that acetylcholine receptors can be viewed as having several biological functions, concluding that the claims fail to stipulate what biological function is intended.

The Office Action also alleges that Claims 42, 43 and 45 are unclear in that the metes and bounds of the invention are unclear. It adds that Claims 42, 43 and 45 are indefinite because they recite "the sequence between position 43 and 1368 of SEQ ID NO:1" without antecedent basis. The Office Action adds that there are many sequences between positions 43 and 1368 of SEQ ID NO:1, and the Office Action alleges that it is unclear to which sequence Applicants refer. The Office Action most helpfully adds that substitution of the phrase "the sequence comprising positions 43-1368 of SEQ ID NO:1" for the phrase "the sequence between position 43 and 1368 of SEQ ID NO:1" would be acceptable.

Claims 42, 43 and 45 are rejected under 35 U.S.C. Section 112, second paragraph as indefinite on the basis that it is allegedly unclear what is intended by a "70% identity" or a "40% identity". The Office Action adds that the specification teaches that preferably identity is calculated using of GCG program GAP, version 10.0, using the "standard settings". The Office Action adds that one of skill in the art would appreciate that there are other available programs which may be used, and that there are a variety of user-chosen parameters which may be varied, and that the calculated identity will vary with the program and parameter values chosen. Because the claims fail to set forth the program and parameter values used in the calculation, the Office Action alleges that one cannot know what is intended by "70% identical" or "40% identical".

By way of response, Applicants have cancelled Claim 45 and provide newly amended Claims 42 and 43 which Applicants believe overcome each of the grounds asserted for rejection above.

Applicants add that with respect to the nucleic acids that encode variants of SEQ ID NO:2, and the allegation that the specification fails to teach which variants will provide a functional beta subunit and which will not, Claims 42 and 43 as amended state that the sequence can be selected from a group which includes sequences which encode a polypeptide having an amino acid sequence which has at least 40% identity to the amino acid sequence as set forth in SEQ. ID NO:2 over its entire length. That this sufficiently defines the invention may be supported by an article such as Peer, Mittl., "Protein Structure Prediction", Biochemisches Institut, Iniversitat Zurich (copy enclosed as Attachment A). From the 10th and 11th page of this reference, it can be inferred that proteins are presumed to have a similar structure if the amino acid sequence identity is >25% over an alignment of at least 80 amino acids. Of curse, proteins having a similar structure would be expected to have a similar function.

Applicants add that with regard to the allegation in the Office Action that because the claims fail to set forth the program and parameter values used in the GCG program GAP, version 10.0, calculation of identity, the Office Action alleges that one cannot know what is intended by "70% identical" or "40% identical", Applicants respectfully traverse and assert this is not well founded. Applicants

respectfully assert that the present invention's specification is much more clear than that of issued patents that are not nearly as precise. First, those skilled in the art would understand that most of the user settable parameters of the software do not affect the calculation of identity. Second, those skilled in the art would understand that the software is provided with "presettings". This is meant to be standard settings that arrive with the software. Therefore, using "standard settings" in line with the specification cannot be interpreted to interpreted to mean that the percentage of identity is changed by setting user-adjustable parameters. Those skilled in the art would understand what is meant by the term "standard settings" and those skilled in the art would all use the same standard settings to arrive at the same percentage of identity.

Applicants submit that the instant application is in condition for allowance. Accordingly, early examination and a Notice of Allowance are respectfully requested for Claims 42 and 43. If the Examiner is of the opinion that the instant application is in condition for other than allowance, he is requested to contact the applicants' Attorney at the telephone number given below so that additional changes may be discussed.

Respectfully submitted,

By Raymond J. Harmuth

Raymond J. Harmuth
Attorney for Applicants
Reg. No. 33,896

Bayer Corporation
100 Bayer Road
Pittsburgh, Pennsylvania 15205-9741
PHONE: (412) 777-8366
FACSIMILE PHONE NUMBER:
412-777-8363
s/rmc/rjh/0097

VERSION WITH MARKINGS TO SHOW CHANGES MADE.

IN THE CLAIMS:

Please cancel Claim 45, and amend Claims 42 and 43 as follows:

42. (Twice Amended) A method of determining a compound which alters ~~at least one~~the conductive property of an acetylcholine receptor, said acetylcholine receptor comprising a polypeptide encoded by a nucleic acid ~~having~~comprising a sequence selected from the group consisting of a sequence as set forth in SEQ ID NO: 1, ~~subsequences of SEQ ID NO: 1 which are at least 14 base pairs in length, sequences which hybridize with SEQ ID NO: 1, sequences which have at least 70% identity to the sequence between position 43 and position 1368 of SEQ ID NO: 1, sequences which are complementary to SEQ ID NO: 1, and~~ sequences which, owing to the degeneracy of the genetic code, encode the same amino acid sequence as ~~do the sequences defined above~~does the sequence as set forth in SEQ ID NO:1, ~~or alters at least one property of a polypeptide exerting the biological function of an acetylcholine receptor ß subunit and comprising an amino acid sequence having at least 40% identity to SEQ ID NO: 2, the compound useful for crop protection and/or pharmaceutical treatment of humans,~~ and sequences which encode a polypeptide having an amino acid sequence which has at least 40% identity to an amino acid sequence as set forth in SEQ ID NO:2 over its entire length

the method comprising:

culturing in the presence of the at least one compound a host cell stably transfected or transformed with a nucleic acid comprising a sequence selected from the group consisting of a sequence as set forth in SEQ ID NO: 1, ~~subsequences of SEQ ID NO: 1 which are at least 14 base pairs in length, sequences which hybridize with SEQ ID NO: 1, sequences which have at least 70% identity to the sequence between position 43 and position 1368 of SEQ ID NO: 1, sequences which are complementary to SEQ ID NO: 1, and~~ sequences which, owing to the degeneracy of the genetic code, encode the same amino acid sequence as ~~do the sequences defined above or a vector comprising an~~

~~isolated and purified nucleic acid molecule as defined above~~ does the sequence as set forth in SEQ ID NO:1, and sequences which encode a polypeptide having an amino acid sequence which has at least 40% identity to the amino acid sequence as set forth in SEQ ID NO: 2 over the entire length, and

detecting the ~~at least one~~ altered conductive property of the receptor.

43. (Twice Amended) A method of determining a compound specifically binding to an acetylcholine receptor which compound upon binding alters the conductive property of the acetylcholine receptor, said acetylcholine receptor comprising a polypeptide encoded by a nucleic acid comprising a sequence selected from the group consisting of a sequence as set forth in SEQ ID NO: 1, ~~subsequences of SEQ ID NO: 1 which are at least 14 base pairs in length, sequences which hybridize with SEQ ID NO: 1, sequences which have at least 70% identity to the sequence between position 43 and position 1368 of SEQ ID NO: 1, sequences which are complementary to SEQ ID NO: 1, and~~ sequences which, owing to the degeneracy of the genetic code, encode the same amino acid sequence as ~~do the sequences defined above, or a polypeptide exerting the biological function of an acetylcholine receptor ß subunit and comprising an amino acid sequence having at least 40% identity to SEQ ID NO:2~~ does the sequence as set forth in SEQ ID NO:1, and sequences which encode a polypeptide having an amino acid sequence which has at least 40% identity to the amino acid sequence as set forth in SEQ ID NO: 2 over the entire length,

the method comprising:

exposing a host cell stably transfected or transformed with a nucleic acid comprising a sequence selected from the group consisting of a sequence as set forth in SEQ ID NO: 1, ~~subsequences of SEQ ID NO: 1 which are at least 14 base pairs in length, sequences which hybridize with SEQ ID NO: 1, sequences~~

~~which have at least 70% identity to the sequence between position 43 and position 1368 of SEQ ID NO: 1, sequences which are complementary to SEQ ID NO: 1, and~~ sequences which, owing to the degeneracy of the genetic code, encode the same amino acid sequence as ~~do the sequences defined above or a vector comprising an isolated and purified nucleic acid molecule as defined above~~ does the sequence as set forth in SEQ ID NO: 1, and sequences which encode a polypeptide having an amino acid sequence which has at least 40% identity to the amino acid sequence as set forth in SEQ. ID NO: 2 over the entire length,

or

exposing a polypeptide encoded by a nucleic acid comprising a sequence selected from the group consisting of a sequence as set forth in SEQ ID NO: 1, ~~subsequences of SEQ ID NO: 1 which are at least 14 base pairs in length, sequences which hybridize with SEQ ID NO: 1, sequences which have at least 70% identity to the sequence between position 43 and position 1368 of SEQ ID NO: 1, sequences which are complementary to SEQ ID NO: 1, and~~ sequences which, owing to the degeneracy of the genetic code, encode the same amino acid sequence as ~~do the sequences defined above or a polypeptide exerting the biological function of an acetylcholine receptor ß subunit and comprising an amino acid sequence having at least 40% identity to SEQ ID NO: 2~~ does the sequence as set forth in SEQ ID NO: 1, and sequences which encode a polypeptide having an amino acid sequence which has at least 40% identity to the amino acid sequence as set forth in SEQ. ID NO: 2 over the entire length,,

or

exposing an acetylcholine receptor comprising a polypeptide encoded by a nucleic acid comprising a sequence selected from the group consisting of a sequence as set forth in SEQ ID NO: 1, ~~subsequences of SEQ ID NO: 1 which are at least 14 base pairs in length, sequences which hybridize with SEQ ID NO: 1, sequences which have at least 70% identity to the sequence between position 43 and position 1368 of SEQ ID NO: 1, sequences which are complementary to SEQ ID NO: 1, and~~ sequences which, owing to the degeneracy of the

genetic code, encode the same amino acid sequence as ~~do the sequences defined above or an acetylcholine receptor comprising a polypeptide exerting the biological function of an acetylcholine receptor ß subunit and comprising an amino acid sequence having at least 40% identity to SEQ ID NO: 2~~ does the sequence as set forth in SEQ ID NO: 1, and sequences which encode a polypeptide having an amino acid sequence which has at least 40% identity to the amino acid sequence as set forth in SEQ. ID NO: 2 over the entire length,

to at least one compound under ~~at least one~~ conditions permitting the interaction of the at least one compound with the host cell, the polypeptide or the receptor, and

identifying the compound specifically binding to the receptor.

ATTACHMENT A

Protein structure prediction

Peer Mittl, Dr.
Biochemisches Institut, Universität Zürich
Tel: 01-635 6559
e-mail: mittl@bioc.unizh.ch

Overview

1. Introduction
- Protein structure prediction - why do we need this?
- Principles of protein structure.
  - the conformational space
  - hierarchical organisation of protein structures
- Sequence/structure relationship
- Protein structure databanks

2. Disciplines of structure prediction
- Secondary structure prediction
  - The Chou & Fassman method (examples)
  - Neuronal networks (examples)
- *ab initio* three-dimensional structure prediction
- Threading (fold-recognition methods)
  - 1D-3D profiles
  - knowledge based potentials
- Homology modelling

Protein structure prediction - why do we need this?

1. Folding & structure prediction are two closely related issues

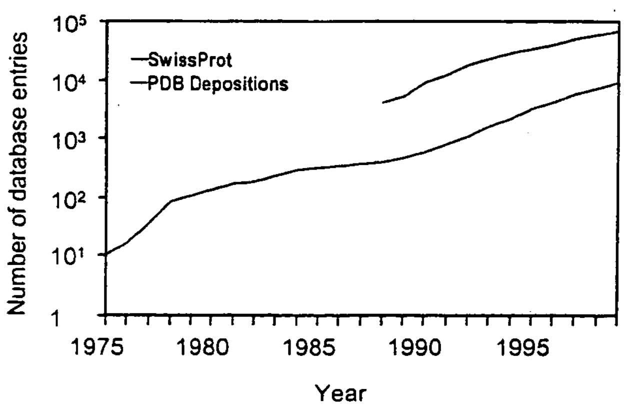
native ribonuclease	$\xrightarrow{\text{heat, deduce}}$	unfolded ribonuclease	$\xrightarrow{\text{oxidise}}$	refolded ribonuclease
100 % activity		no activity		fully active
well defined 3D structure		random coil structure		well defined 3D structure

==> Information for structure is hidden in the amino acid sequence.

2. Functional aspects:
   - functional details (which residue is involved in ...)
   - function for novel proteins (genome sequencing projects)
   - epitopes for protein/protein interactions (antigen/antibodies)
   - drug binding (resistance against anti-HIV drugs)

   ....

3. Number of experimentally determined 3D-structures lacks behind the number of protein sequences.

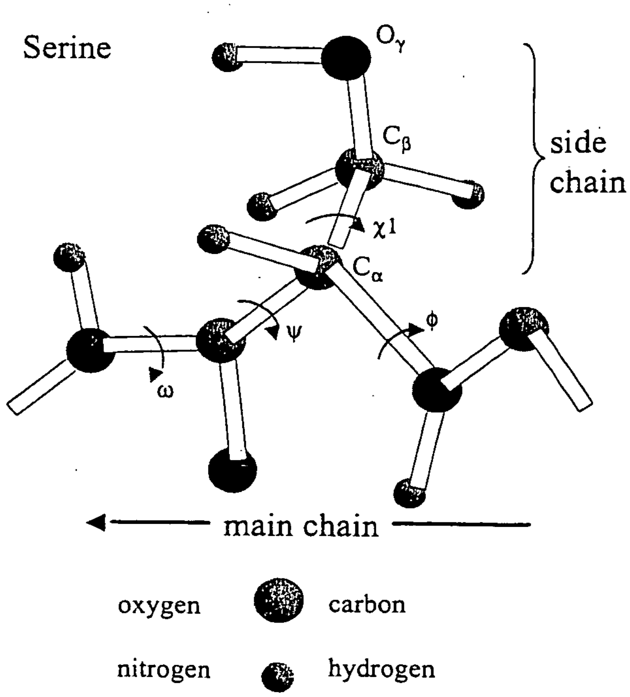


Principles of Protein Structure

Protein structures are organised hierarchically.

level of hierarchy	elements
atoms	H, C, N, O, S (P, Se)
residues	20 natural amino acids (Ala, ... Trp), modified amino acids
secondary structure	α-helix, β-sheet (parallel/anti-parallel), loops, 3_{10}-helix, collagen helix
super-secondary structure	4-helix bundle fold (α–α–α–α), Rossman fold (α–β–α–β), ...
domains	Ig-like domain, NAD-binding domain, ...
globular proteins	myoglobin, haemoglobin, actin, ...
protein aggregates	homo/hetreo dimers, trimers, ... , viruses, filaments, (ribo)somes

Atoms and residues

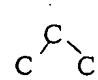


Conformation defined by:

1. bond length (2 atoms)

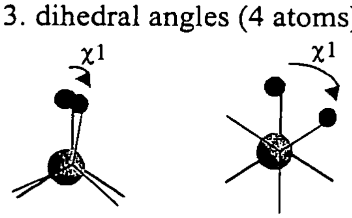
C----C	1.50 Å ± 0.03Å
C----OH	1.35 Å ± 0.03Å
C----NH_2	1.35 Å ± 0.03Å
C==O	1.25 Å ± 0.03Å

2. bond angles (3 atoms)



106° ± 3°

3. dihedral angles (4 atoms)

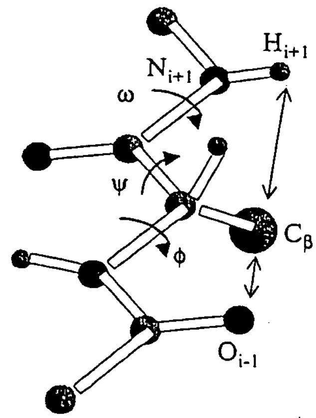


ecliptic	staggered
χ1=0°, 120°, 240°	χ1=60°, 180°, 300°
± 20°	± 40°
unfavourable	favourable

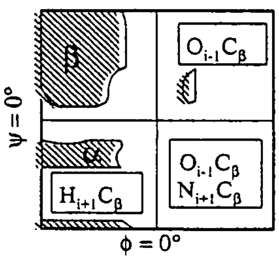
Dihedral angles of the main-chain

The conformation of the main-chain is restricted:

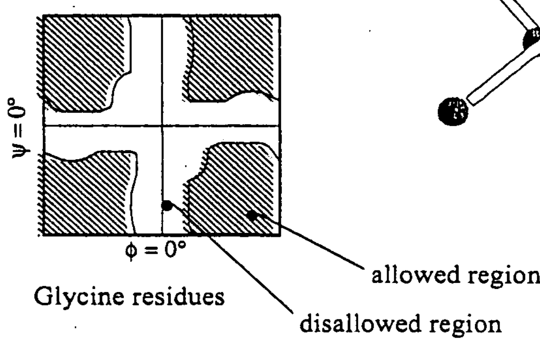
1. peptide bond mainly *trans* ($\omega = 180°$),
   cis conformations ($\omega = 0°$) before proline residues
2. ϕ/ψ–angles are restricted due to clashes between C_β-atom and main-chain atoms



ϕ/ψ–plots (Ramachandran plot)



non-Glycine residues



Glycine residues

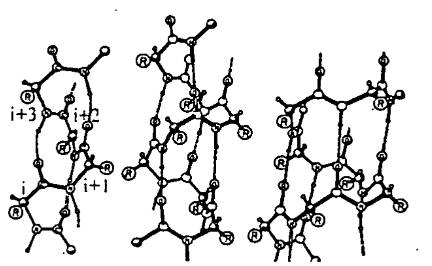
Secondary structural elements

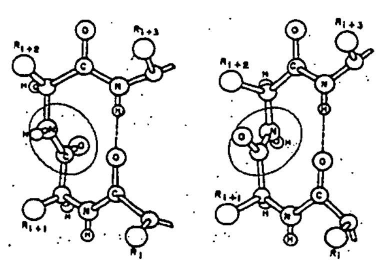
Secondary structural elements

repetitive elements:	α-helix, β-sheet (parallel/anti-parallel), 3_{10}-helix, collagen helix
loops:	reverse turns (several types), random coil

Secondary structural elements are defined by hydrogen bonds

3_{10}-helix $O_i \ldots N_{(i+3)}$ | α-helix $O_i \ldots N_{(i+4)}$ | (π-helix) $O_i \ldots N_{(i+5)}$

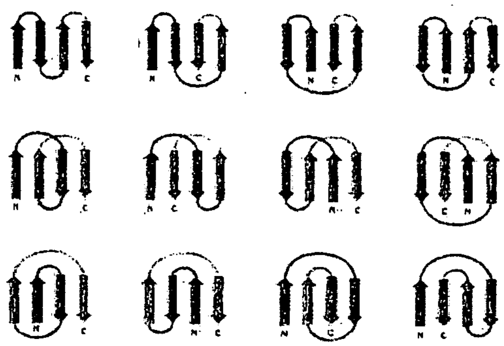




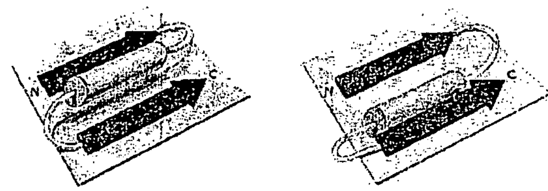
Two types of reversed turns

Super-secondary structure

- Secondary structural elements are combined to form higher structural aggregates (fold, motifs).
- Defined by: secondary structural elements involved
  relative orientation between elements
  connectivity between elements
- Three main classes: all-α, all-β, αβ



12 different all-β folds

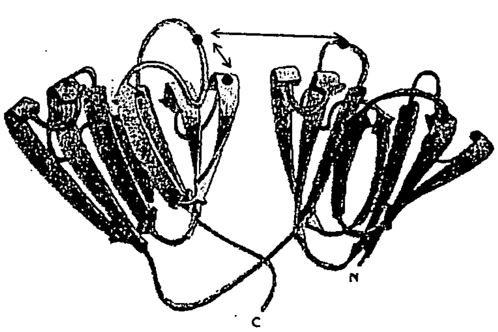


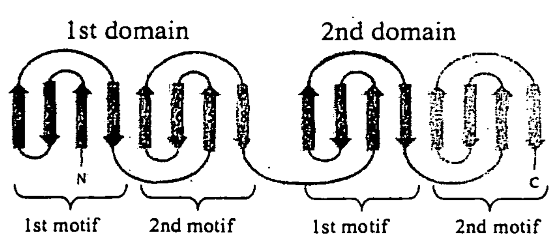
2 different αβ folds

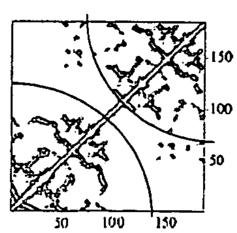
Domains

- domains consist of super-secondary structural elements
- residues that are fare apart in the sequence but close together in space form a domain (neighborhood relationship)
- domains are detected by Ca-distance matrix
- little flexibility within a domain but large movements between domains

Example: gamma-crystallin



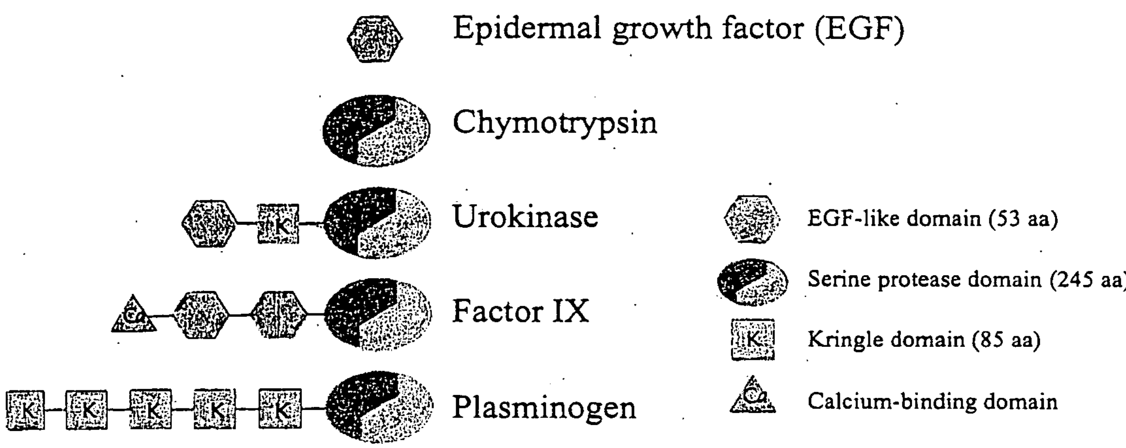




Cα-distance matrix
red: short distance
blue: longer distance
white: very long distance

Globular Proteins

Domains are modular building blocks for globular proteins



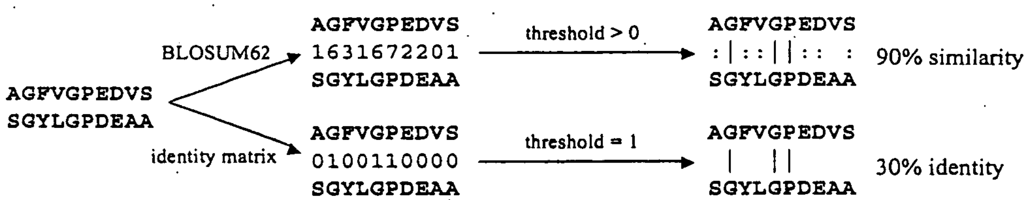
Protein aggregates

- globular proteins aggregate to form higher order oligomers
- oligomerization is guided by complementary surfaces (shape, electrostatic properties, hydrophobic patches, ...)

Stöchiometry	Symmetry		
	no sym.	rotational sym.	helical sym.
A_1	myoglobin	-	-
....			
A_4	-	haemoglobin	-
A_2B_2	-	caspase-3	-
.....			
A,B,C,.....	ribosome	picorna virus	-
A_{99999}	amorphous	-	actin

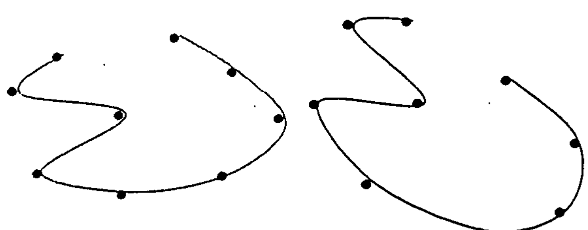
2. Sequence / structure relationship

- Methods of sequence/sequence comparison (lecture 8 & 9, G. Capitani)
  different alignment methods, optimisation of similarity score
  sequence similarity: substitution matrix, threshold value
  sequence identity: identity matrix, threshold=1



- Methods of structure/structure comparison
  1.) the root mean square deviation (rmsd)
  2.) secondary structure identity

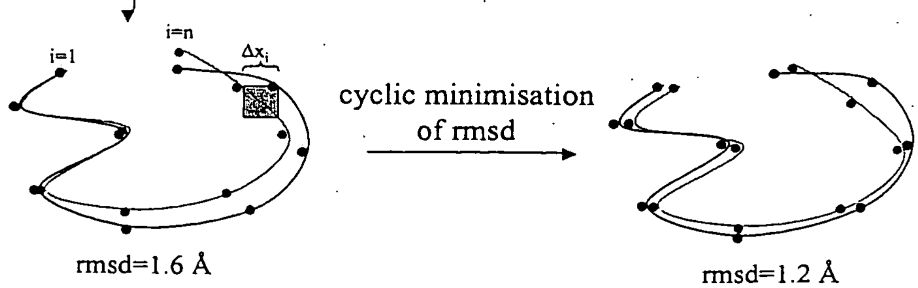
Root mean square deviation (rmsd)



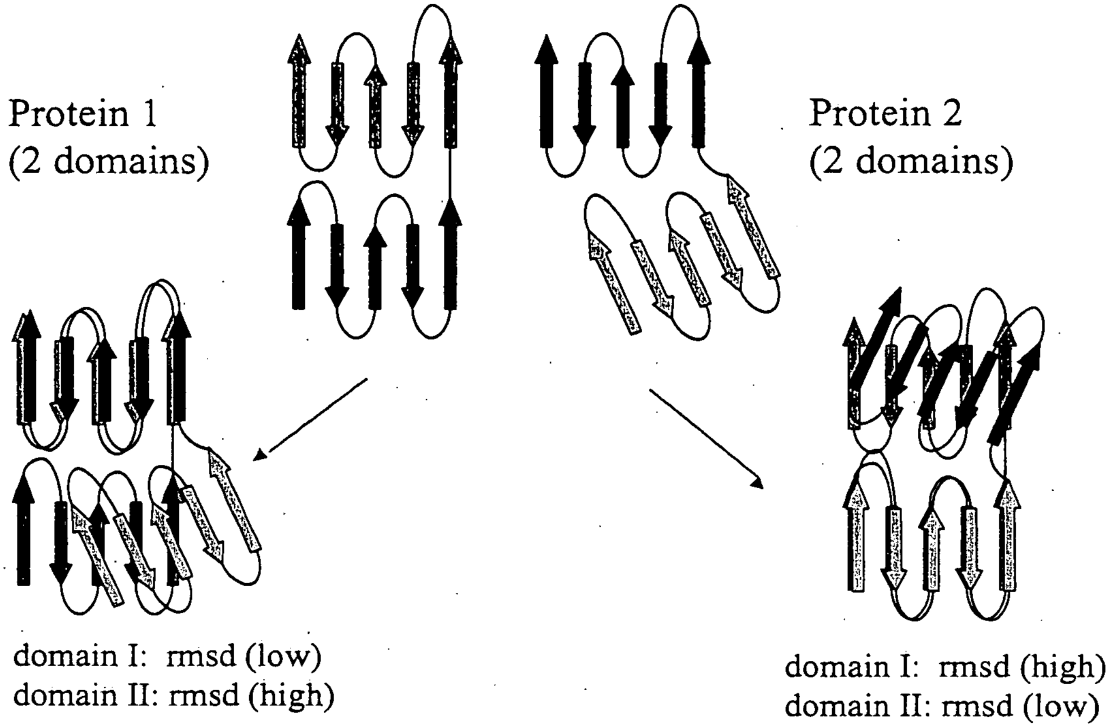
$$\text{rmsd} = \sqrt{\frac{\sum_{i=1}^{n} (\Delta \mathbf{x}_i)^2}{n}}$$

for atom pairs i=1 to n
$\Delta \mathbf{x}_i$ distance between two compared atoms in space

find starting superposition



Rigid body movements increase the *rmsd*



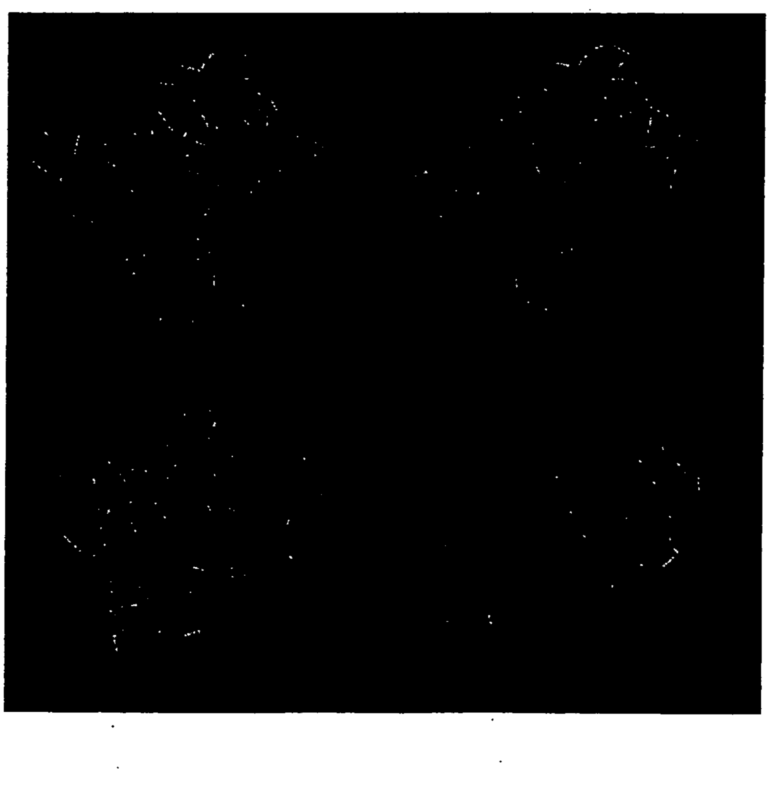
Sequence identity and rmsd of Sperm Whale myoglobin

myoglobin
pig

rmsd = 0.5 Å
id = 86%

haemoglobin
pig

rmsd = 1.5 Å
id = 28%



globin-3
P. piclitum

rmsd = 2.2 Å
id = 18%

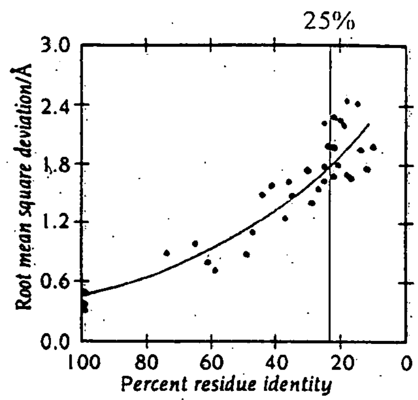
phycocyanin
F. diplosiphon

rmsd = 3.3 Å
id = 8%

How does sequence identity correlate with structural similarity?

Analysis by Chotia & Lesk (1986)

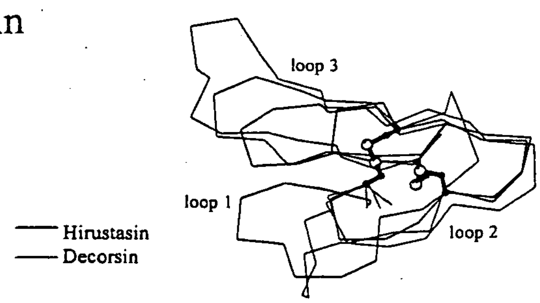
- 100 % sequence identity: rmsd is at the order of the experimental co-ordinate error
- < 25 % sequence identity: structures might be similar (twilight-zone), but they might also be different
- rigid body movement can enhance rmsd although structures are similar



Sequence / sequence *versus* structure / structure alignment

Example: Hirustasin and Decorsin are leech derived inhibitors

25.6 % sequence identity

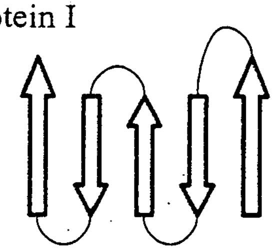



```
Decorsin (sequence)   ---------------APRLPQCQGDDQE---------KCLCNK----DECPPGQCRFPRGDADP-YCE----
                                     :     |:: :           ||:||:     :|· |  :  :|   |  |
Hirustasin            -----------TQGNTCGGETCSAAQVCLK------GKCVCNEVHCRIRCKYGLKKDENGCEYPCSCAKASQ
                                 .:   |. : |     |          |
Decorsin (structure)  APRLPQCQGDDQEKCLCNKDECPP-GQCRFPRGDADPYCE--------------------------------
                                      loop1 loop2  loop3
```

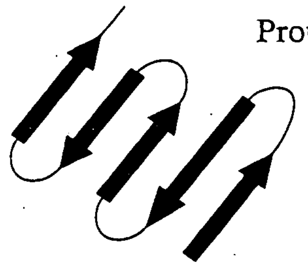
Conclusion: at low sequence identity the sequence/sequence alignment might imply a wrong structure/structure alignment.

Secondary structure identity as a measure for structural identity

Protein I

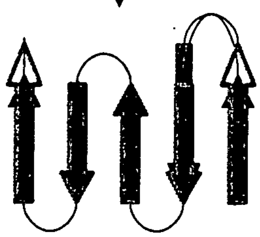


Protein II



1. align sequences
2. assign secondary structure to sequence

superimpose structures



```
1           10          20          30     37
BBBBBBCCCCBBBBCCCBBBBCCCBBBBCCCBBBBB
BBBBCCCCBBBBBBCCCBBBBCCCBBBBCCCBBBBC
****--**--*************************-
```

(37-5)/ 37 = 0.86

86 % secondary structure identity

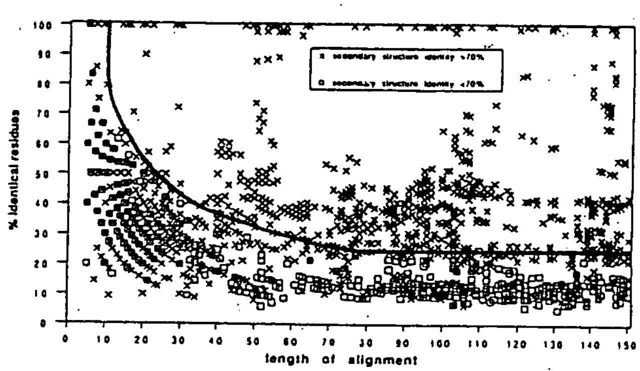
secondary structure identity > 70% ⟶ structures similar

What is the minimal alignment length to deduce structural similarity from sequence identity?

Analysis of Schneider & Sander (1991)

Definition: two structures are similar if:

1. rmsd < 2.5 Å
2. secondary structure identity > 70%



Alignment length	sequence identity threshold
< 10	79.6 %
20	53.9 %
30	43.0 %
40	36.6 %
60	29.1 %
> 80	24.8 %

Conclusion

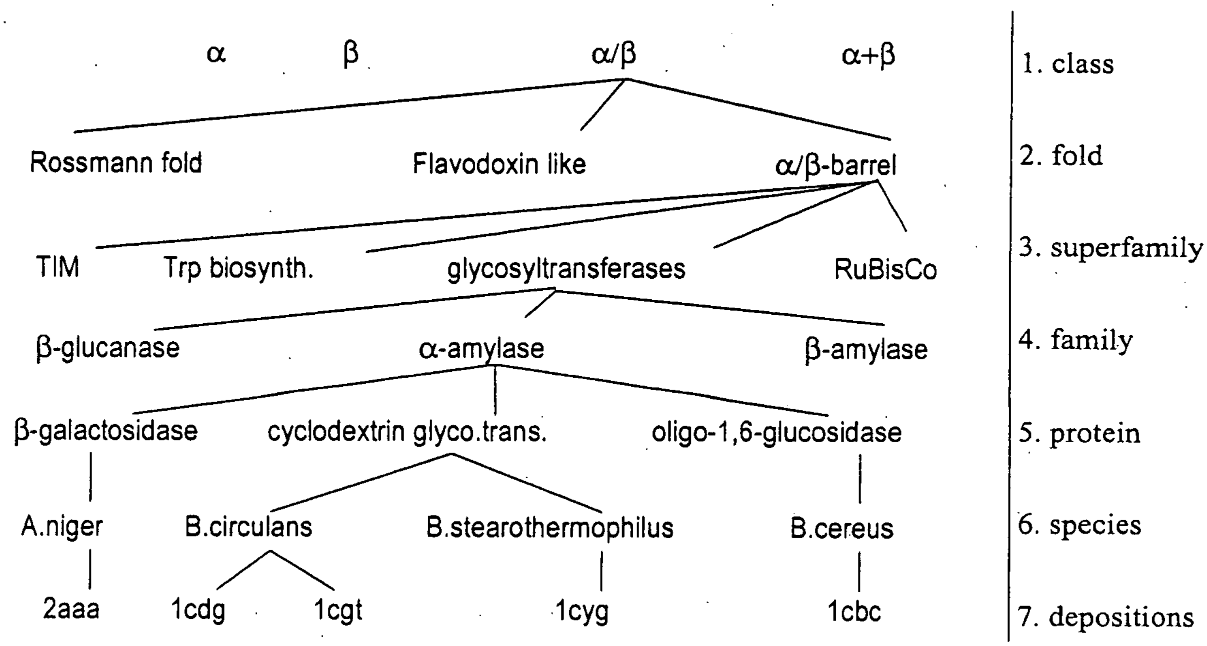
- We can deduce structural similarity between two proteins if the sequence identity is > 25 % over an alignment length of at least 80 amino acids.
- Protein structures can be similar even at very low sequence identity (twilight-zone).
- Structure is better conserved than sequence.

Databanks for protein structures

Brookhaven Protein databank (PDB)	experimentally determined protein structures http://www.rcsb.org/pdb/
Cambridge Crystallographic Data Centre	Crystal structures of small molecules http://www.ccdc.cam.ac.uk/
BioMagResBank	NMR protein structures http://www.bmrb.wisc.edu/
ModBase	Database of homology models http://pipe.rockefeller.edu/
SCOPE	hierarchical clustering of structures http://scop.mrc-lmb.cam.ac.uk/scop/
CATH	" http://www.biochem.ucl.ac.uk/bsm/cath_new
FSSP	" http://www2.embl-ebi.ac.uk/dali/fssp/fssp.html
Relibase	structures of proteins/small molecule complexes http://relibase.ebi.ac.uk/

Hierarchical Structure Classifications in SCOP

Murzin, Brenner, Hubbard & Chothia; 1995



Structural Classificationof Proteins

Superfamily: Globin-like

Lineage:

1. Root: scop
2. Class: All alpha proteins
3. Fold: Globin-like
   core:6 helices;foldedeaf,partly opened
4. Superfamily: Globin-like

Families:

1. Truncated hemoglobin (2)
2. Globins (52)
   Heme-bindingprotein
3. Phycocyanins (10)
   oligomersof two differenttypesof homologousubunits
   each subunit contains 2 additional helices at the N-terminus
   bindsa chromophore

Protein Domains:

1. Hemoglobin I
   1. Ark clam (*Scapharcinaequivalvis*) (10)
   2. Clam (*Lucina pectinata*) (4)
2. Glycera globin
   1. Marine bloodworm (*Glyceradibranchiata*) (4)
3. Myoglobin
   1. Sperm whale (*Physetercatodon*) (131)
   2. Sea hare (*Aplysia limacina*) (7)
   3. Common seal (*Phocavitulina*) (1)
   4. Pig (*Sus scrofa*) (17)
   5. Horse (*Equuscaballus*) (13)

   ⋮

PDB Entry Domains:

1. 1a6m
   complexedwith hem,oxy, so4
2. 1a6k
   complexedwith hem,so4
3. 1bzp
   complexedwith hem,so4
   1. chain a

⋮

Disciplines of structure prediction

1.) Prediction of secondary structure
 a. method of Chou & Fassman
 b. neural networks

2.) Prediction of tertiary structure
 a. *ab initio* structure prediction
 b. threading
  -1D-3D profiles
  - knowledge based potentials
 c. homology modelling

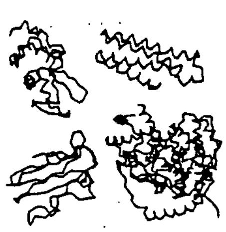
predictive methods

modelling methods

The Chou & Fassman method for secondary structure prediction

Chou & Fassman (1974) *Biochemistry* **13**, 211-221, 222-245

1. Probabilities for all amino acids to be either in α-helix, β-sheet or coil.
   20 * 3 probabilities
2. Set knowledge-based rules to apply probability tables for prediction



Databank of 15 structures

protein 1 AGPDFVILKRW...
CCHHHHCCHHH...

protein 2 PGSAALHVIYI...
CCBBBBCCHHH...

⋮

protein 15 MVEDEHVILTAGF...
CBBBBBBBBBHHH...

Assignment of secondary structures to sequences

Amino acid	No. of residues $\#_{total}$	$\#_{helix}$	$\#_{sheet}$	$\#_{coil}$
Ala	228	119	38	71
Arg	78	22	12	44
......	...	...	...	...
Val	181	74	51	56

List of observations

From observations to probabilities

Aminoacid	$\#_{total}$	observations $\#_{helix}$	$\#_{sheet}$	$\#_{coil}$		frequency f_{helix}	f_{sheet}	f_{coil}		probability P_{helix}	P_{sheet}	P_{coil}
Ala	228	119	38	71		0.52	0.97	0.31		1.45	0.97	0.66
Arg	78	22	12	44		0.28	0.15	0.56		0.79	0.90	1.20
......	...	...	...	...	→	...	...	...	→	...	...	...
Val	181	74	51	56		0.41	0.28	0.31		1.14	1.65	0.66
					average $\langle f \rangle$	0.36	0.17	0.47		1.0	1.0	1.0

$119 / 228 = 0.52$

$\#^{Ala}_{helix} / \#^{Ala}_{total} = f^{Ala}_{helix}$

$0.52 / 0.36 = 1.45$

$f^{Ala}_{helix} / \langle f_{helix} \rangle = P^{Ala}_{helix}$

Knowledge based rules for the prediction of secondary structural elements

1. Method predicts α-helix, β-sheet and β-turns
2. Rules for secondary structure assignment:
   a. Find the nucleation centre for α-helix or β-sheet
      α-helix: 4 residues out of 6 with $P_{helix} > 1.06$
      $\langle P_{helix} \rangle_6 > 1.03$
      β-sheet: 3 residues out of 5 with $P_{sheet} > 1.05$
   b. Extend nucleation centre to both sides
      α-helix: stop if $\langle P_{helix} \rangle_4 < 1.00$ for 4 residues
      β-sheet: stop if $\langle P_{sheet} \rangle_4 < 1.00$ for 4 residues
   c. Decide between α-helix and β-sheet
      α-helix: $\langle P_{helix} \rangle_4 > \langle P_{sheet} \rangle_4$
      β-sheet: $\langle P_{sheet} \rangle_4 > \langle P_{helix} \rangle_4$

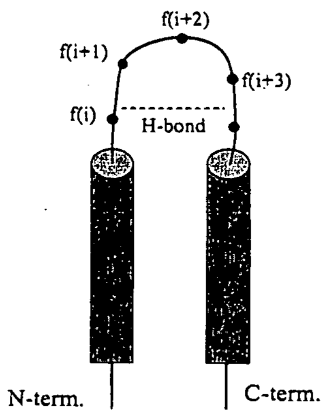
Prediction of β-turns

β-turns: $<P_{turn}>_4$ for 4 residues > 1.00

$<P_{turn}>_4 \; > \; <P_{helix}>_4$ or $<P_{sheet}>_4$

$p_{turn} = f_i * f_{i+1} * f_{i+2} * f_{i+3}$,

$p_{turn} > 0.75 * 10^{-4}$



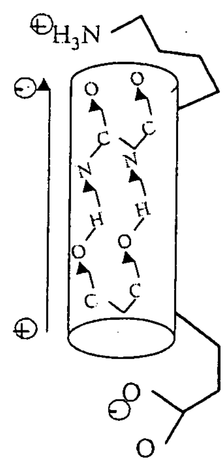
β-turn frequency

	f(i)	f(i+1)	f(i+2)	f(i+3)
A	0.060	0.076	0.035	0.058
C	0.149	0.053	0.117	0.128
D	0.147	0.110	0.179	0.081
E	0.056	0.060	0.077	0.064
F	0.059	0.041	0.065	0.065
G	0.102	0.085	0.190	0.152
H	0.140	0.047	0.093	0.054
I	0.043	0.034	0.013	0.056
K	0.055	0.115	0.072	0.095
L	0.061	0.025	0.036	0.070
M	0.068	0.082	0.014	0.055
N	0.161	0.083	0.191	0.091
P	0.102	0.301	0.034	0.068
Q	0.074	0.098	0.307	0.098
R	0.070	0.106	0.099	0.085
S	0.120	0.139	0.125	0.106
T	0.086	0.108	0.065	0.079
V	0.062	0.048	0.028	0.053
W	0.077	0.013	0.064	0.167
Y	0.082	0.065	0.114	0.125

Amino acid preference for α-helices

TABLE IV: Frequency of Helical Boundary and Central Residues[a] in 15 Proteins.

	f_{bN}[b]		f_{bC}[c]		f_{abN}[d]		f_{abC}[e]		f_{h1}[f]
Pro	0.212	His(+)	0.216	His(+)	0.162	His(+)	0.149	Ala	0.184
Asp(−)	0.207	Lys(+)	0.160	Pro	0.141	Asp(−)	0.135	Phe	0.183
Glu(−)	0.195	Gln	0.158	Ser	0.104	Lys(+)	0.120	Leu	0.179
Ala	0.140	Arg(+)	0.154	Gly	0.103	Asn	0.120	Glu(−)	0.177
Trp	0.136	Cys	0.148	Asn	0.098	Arg(+)	0.115	Val	0.166
Thr	0.122	Met	0.143	Ile	0.085	Gly	0.112	Gln	0.158
Gln	0.116	Glu(−)	0.124	Phe	0.085	Ile	0.094	Met	0.143
Phe	0.098	Ala	0.118	Gln	0.084	Pro	0.094	Lys(+)	0.126
Asn	0.090	Val	0.116	Leu	0.082	Cys	0.093	Trp	0.114
Ser	0.079	Phe	0.110	Asp(−)	0.081	Thr	0.090	Asp(−)	0.099
Cys	0.074	Leu	0.102	Glu(−)	0.080	Tyr	0.090	Cys	0.093
Met	0.071	Asn	0.090	Tyr	0.080	Phe	0.073	Arg(+)	0.090
Tyr	0.070	Ser	0.084	Lys(+)	0.074	Met	0.071	Ser	0.084
Ile	0.066	Ile	0.075	Val	0.072	Leu	0.061	Thr	0.083
Val	0.061	Asp(−)	0.054	Met	0.071	Glu(−)	0.053	Asn	0.075
Gly	0.060	Tyr	0.050	Trp	0.068	Ala	0.044	Ile	0.075
Lys(+)	0.057	Thr	0.045	Thr	0.064	Gln	0.042	His(+)	0.068
Leu	0.056	Trp	0.045	Cys	0.056	Ser	0.040	Tyr	0.050
His(+)	0.054	Gly	0.039	Ala	0.044	Trp	0.023	Gly	0.034
Arg(+)	0.038	Pro	0	Arg(+)	0.038	Val	0.022	Pro	0
$\langle f \rangle$[g]	0.097		0.097		0.082		0.080		0.112



- proline residues frequently at N-terminus, never at C-terminus or inside an a-helix.
- negatively charged side-chains at N-terminus.
- positively charged side-chains at C-terminus, rarely at N-terminus.

4. Expand nucleation site for β-sheet

1.02

1.04

1.06

1.08

1.33 1.37 1.10

1.45	1.51	1.51	1.16	1.21	1.16	1.16	0.77	1.16	1.08	1.08	1.13
Met -	Glu -	Glu -	Lys -	Leu -	Lys -	Lys -	Ser -	Lys -	Ile -	Ile -	Phe
1.05	0.37	0.37	0.74	1.30	0.74	0.74	0.75	0.74	1.60	1.60	1.38

1.33

1.17

0.96

1.06	1.06	0.57	0.57	0.57	0.57	0.77	0.57	1.16	0.57	0.83	1.11
Val -	Val -	Gly -	Gly -	Pro -	Gly -	Ser -	Gly -	Lys -	Gly -	Thr -	Gln
1.70	1.70	0.75	0.75	0.55	0.75	0.75	0.75	0.74	0.75	1.19	1.10

1.23

0.97

6. Expand the β-turn nucleation sites

Met - Glu - Glu - Lys - Leu - Lys - Lys - Ser - Lys - Ile - Ile - Phe

Val - Val - Gly - Gly - Pro - Gly - Ser - Gly - Lys - Gly - Thr - Gln

1.55 1.13

0.50	0.50	1.56	1.56	1.52	1.56	1.43	1.56	1.01	1.56	0.96	0.98
	0.45	0.102	0.085	0.034	0.152						
	6.18		0.102	0.301	0.190	0.106					
	1.65			0.102	0.085	0.125	0.152				
	2.56				0.102	0.139	0.190	0.095			
	1.12					0.120	0.085	0.072	0.152		
	1.76						0.102	0.115	0.190	0.079	
	0.30							0.055	0.085	0.065	0.098

SEQUENCE LISTING

<110> Bayer Aktiengesellschaft

<120> Nucleic acids coding for new acetylcholine receptor beta subunits of insects

<130> Le A 34 147

<140>
<141>

<150> DE 199 59 582.8
<151> 1999-12-10

<160> 4

<170> PatentIn Ver. 2.1

<210> 1
<211> 1539
<212> DNA
<213> Drosophila melanogaster

<220>
<221> CDS
<222> (43)..(1365)

<400> 1
attcggcacg agggtacatc cgaaacaaag gcgcgctgaa ca atg acg acg act 54
Met Thr Thr Thr
1

ccc aag ata aag gca cca gtt tcc ggt cct gga ctg cca cta ctg ctg 102
Pro Lys Ile Lys Ala Pro Val Ser Gly Pro Gly Leu Pro Leu Leu Leu
5 10 15 20

caa atg cta atg ggg atg ctt ctt atg ggg ctg act tcc gtg cca ggc 150
Gln Met Leu Met Gly Met Leu Leu Met Gly Leu Thr Ser Val Pro Gly
25 30 35

gcc act gcc acc gcg gac ccc aag aac gcc aat gtc aag gcc ctg gat 198
Ala Thr Ala Thr Ala Asp Pro Lys Asn Ala Asn Val Lys Ala Leu Asp
40 45 50

cgc ctc cac gcc ggc ctg ttc acg aac tac gac agc gat gtg cag ccg 246
Arg Leu His Ala Gly Leu Phe Thr Asn Tyr Asp Ser Asp Val Gln Pro
55 60 65

gtg ttc caa gga acc ccc acg aac gtg tcc ctg gaa atg gtg gtc acc 294
Val Phe Gln Gly Thr Pro Thr Asn Val Ser Leu Glu Met Val Val Thr
70 75 80

tac ata gac atc gac gag ttg aac ggc aag ctg acc acc cac tgc tgg 342
Tyr Ile Asp Ile Asp Glu Leu Asn Gly Lys Leu Thr Thr His Cys Trp
85 90 95 100

ctg aat ctc cga tgg aga gac gag gag cgc gtg tgg caa ccg tca caa 390
Leu Asn Leu Arg Trp Arg Asp Glu Glu Arg Val Trp Gln Pro Ser Gln
105 110 115

tat gac aac atc acg cag atc act ttg aag tcc agc gag gtc tgg acc 438
Tyr Asp Asn Ile Thr Gln Ile Thr Leu Lys Ser Ser Glu Val Trp Thr
120 125 130

ccc caa atc aca ctc ttc aac ggc gac gaa ggt ggc ctg atg gcc gaa 486
Pro Gln Ile Thr Leu Phe Asn Gly Asp Glu Gly Gly Leu Met Ala Glu
135 140 145

```
acc cag gtg acc ctc agc cac gat ggc cac ttc cgg tgg atg cct cca      534
Thr Gln Val Thr Leu Ser His Asp Gly His Phe Arg Trp Met Pro Pro
    150                 155                 160

gcc gtg tac acg gcc tac tgc gaa ctc aac atg ctc aac tgg ccc cac      582
Ala Val Tyr Thr Ala Tyr Cys Glu Leu Asn Met Leu Asn Trp Pro His
165                 170                 175                 180

gac aag cag agc tgc aag ttg aag atc ggc tcc tgg ggc ctg aag gtc      630
Asp Lys Gln Ser Cys Lys Leu Lys Ile Gly Ser Trp Gly Leu Lys Val
                185                 190                 195

gtc ctg ccg gag aac ggc acg gcg aga gga gag tcc ctt gac cac gac      678
Val Leu Pro Glu Asn Gly Thr Ala Arg Gly Glu Ser Leu Asp His Asp
            200                 205                 210

gac ctg gtt cag tca ccg gag tgg gaa atc gtg gac tcg cga gcc cac      726
Asp Leu Val Gln Ser Pro Glu Trp Glu Ile Val Asp Ser Arg Ala His
        215                 220                 225

ttt gtc agt cag gac tac tac ggc tac atg gag tac act ctg acg gct      774
Phe Val Ser Gln Asp Tyr Tyr Gly Tyr Met Glu Tyr Thr Leu Thr Ala
    230                 235                 240

cag cgg cgc tcc tcc atg tac acg gcc gtc atc tac aca ccc gcg tcc      822
Gln Arg Arg Ser Ser Met Tyr Thr Ala Val Ile Tyr Thr Pro Ala Ser
245                 250                 255                 260

tgc atc gtc atc ctg gcc ctc tca gcc ttc tgg ctg cct ccc cac atg      870
Cys Ile Val Ile Leu Ala Leu Ser Ala Phe Trp Leu Pro Pro His Met
                265                 270                 275

ggc ggc gag aag atc atg atc aac ggc ctg ctc atc atc gtg atc gcc      918
Gly Gly Glu Lys Ile Met Ile Asn Gly Leu Leu Ile Ile Val Ile Ala
            280                 285                 290

gcc ttc ctc atg tac ttc gcc cag ctc ctg cca gtg ctg tcc aac aat      966
Ala Phe Leu Met Tyr Phe Ala Gln Leu Leu Pro Val Leu Ser Asn Asn
        295                 300                 305

act cca ctt gtg gta atc ttc tac agc acc agc ctg ctg tat ctg agc      1014
Thr Pro Leu Val Val Ile Phe Tyr Ser Thr Ser Leu Leu Tyr Leu Ser
    310                 315                 320

gtc tcc acc atc gtc gag gtt cta gtt ctg tac ctg gcc aca ggc aag      1062
Val Ser Thr Ile Val Glu Val Leu Val Leu Tyr Leu Ala Thr Gly Lys
325                 330                 335                 340

cac aag agg cgc ctg ccg gag gcg ctg aga aag ctg ctg cac ggg cac      1110
His Lys Arg Arg Leu Pro Glu Ala Leu Arg Lys Leu Leu His Gly His
                345                 350                 355

ctg ggc acg tgg ctg ctg ctc tcg gtg ttc agc acc act ggc gag tcg      1158
Leu Gly Thr Trp Leu Leu Leu Ser Val Phe Ser Thr Thr Gly Glu Ser
            360                 365                 370

cag gcg gag aag acc aaa gag atg gac gag cac ccg tac gag gag gcg      1206
Gln Ala Glu Lys Thr Lys Glu Met Asp Glu His Pro Tyr Glu Glu Ala
        375                 380                 385

gac gag cag gag tcc agt ccg ctg ggc atc aac cac acc gag gtg ccg      1254
Asp Glu Gln Glu Ser Ser Pro Leu Gly Ile Asn His Thr Glu Val Pro
    390                 395                 400

ggc gcc aag gcc aac cag ttc gac tgg gcg ctg ctg gcc acc gcc gtg      1302
Gly Ala Lys Ala Asn Gln Phe Asp Trp Ala Leu Leu Ala Thr Ala Val
405                 410                 415                 420
```

gac cgc att tcc ttc gtt tcc ttc agc ctg gcc ttc ctc att ctg gcc 1350
Asp Arg Ile Ser Phe Val Ser Phe Ser Leu Ala Phe Leu Ile Leu Ala
425 430 435

atc agg tgc tcc gtg tagggatgct cgagactcaa ggccacatcc caagccagtg 1405
Ile Arg Cys Ser Val
440

cgcactctga actagttttg catttgcgat ttcatgtatt taatgtgtgt gcgaacttat 1465

aattatttaa tgatgagacc tcgtatggaa taaaggacct ctgccgaatg tctgcttaca 1525

aaaaaaaaaa aaaa 1539

<210> 2
<211> 441
<212> PRT
<213> Drosophila melanogaster

<400> 2
Met Thr Thr Thr Pro Lys Ile Lys Ala Pro Val Ser Gly Pro Gly Leu
1 5 10 15

Pro Leu Leu Leu Gln Met Leu Met Gly Met Leu Leu Met Gly Leu Thr
20 25 30

Ser Val Pro Gly Ala Thr Ala Thr Ala Asp Pro Lys Asn Ala Asn Val
35 40 45

Lys Ala Leu Asp Arg Leu His Ala Gly Leu Phe Thr Asn Tyr Asp Ser
50 55 60

Asp Val Gln Pro Val Phe Gln Gly Thr Pro Thr Asn Val Ser Leu Glu
65 70 75 80

Met Val Val Thr Tyr Ile Asp Ile Asp Glu Leu Asn Gly Lys Leu Thr
85 90 95

Thr His Cys Trp Leu Asn Leu Arg Trp Arg Asp Glu Glu Arg Val Trp
100 105 110

Gln Pro Ser Gln Tyr Asp Asn Ile Thr Gln Ile Thr Leu Lys Ser Ser
115 120 125

Glu Val Trp Thr Pro Gln Ile Thr Leu Phe Asn Gly Asp Glu Gly Gly
130 135 140

Leu Met Ala Glu Thr Gln Val Thr Leu Ser His Asp Gly His Phe Arg
145 150 155 160

Trp Met Pro Pro Ala Val Tyr Thr Ala Tyr Cys Glu Leu Asn Met Leu
165 170 175

Asn Trp Pro His Asp Lys Gln Ser Cys Lys Leu Lys Ile Gly Ser Trp
180 185 190

Gly Leu Lys Val Val Leu Pro Glu Asn Gly Thr Ala Arg Gly Glu Ser
195 200 205

Leu Asp His Asp Asp Leu Val Gln Ser Pro Glu Trp Glu Ile Val Asp
210 215 220

Ser Arg Ala His Phe Val Ser Gln Asp Tyr Tyr Gly Tyr Met Glu Tyr
225 230 235 240

Thr Leu Thr Ala Gln Arg Arg Ser Ser Met Tyr Thr Ala Val Ile Tyr
245 250 255

Thr Pro Ala Ser Cys Ile Val Ile Leu Ala Leu Ser Ala Phe Trp Leu
260 265 270

Pro Pro His Met Gly Gly Glu Lys Ile Met Ile Asn Gly Leu Leu Ile
275 280 285

Ile Val Ile Ala Ala Phe Leu Met Tyr Phe Ala Gln Leu Leu Pro Val
290 295 300

Leu Ser Asn Asn Thr Pro Leu Val Val Ile Phe Tyr Ser Thr Ser Leu
305 310 315 320

Leu Tyr Leu Ser Val Ser Thr Ile Val Glu Val Leu Val Leu Tyr Leu
325 330 335

Ala Thr Gly Lys His Lys Arg Arg Leu Pro Glu Ala Leu Arg Lys Leu
340 345 350

Leu His Gly His Leu Gly Thr Trp Leu Leu Leu Ser Val Phe Ser Thr
355 360 365

Thr Gly Glu Ser Gln Ala Glu Lys Thr Lys Glu Met Asp Glu His Pro
370 375 380

Tyr Glu Glu Ala Asp Glu Gln Glu Ser Ser Pro Leu Gly Ile Asn His
385 390 395 400

Thr Glu Val Pro Gly Ala Lys Ala Asn Gln Phe Asp Trp Ala Leu Leu
405 410 415

Ala Thr Ala Val Asp Arg Ile Ser Phe Val Ser Phe Ser Leu Ala Phe
420 425 430

Leu Ile Leu Ala Ile Arg Cys Ser Val
435 440

<210> 3
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 3
tggcarccit cicartayga 20

<210> 4
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 4
catratytty tciccicccca t 21